R E P O R T R E S U M E S

ED 016 439 FL 000 725

A CLOSE LOOK AT THE IMPORTANCE AND PLACE OF TESTING IN THE FOREIGN LANGUAGE PROGRAM.

BY- BARRUTIA, RICHARD

PUB DATE 67

EDRS PRICE MF-$0.25 HC-$0.76 17P.

DESCRIPTORS- *LANGUAGE TESTS, *SECOND LANGUAGE LEARNING, *STUDENT TESTING, *TEST CONSTRUCTION, *LANGUAGE INSTRUCTION, ACHIEVEMENT TESTS, TEACHING TECHNIQUES, TEST RESULTS, STUDENT MOTIVATION, ACADEMIC APTITUDE, ACADEMIC ACHIEVEMENT, DIAGNOSTIC TESTS, INSTRUCTIONAL MATERIALS, PACING, TESTING, TESTING PROBLEMS,

TESTING IS ESPECIALLY IMPORTANT IN FOREIGN LANGUAGE PROGRAMS BECAUSE THE NATURE OF AUDIOLINGUAL TEACHING REQUIRES GREAT AMOUNTS OF FEEDBACK, REINFORCEMENT, AND VERIFICATION. TESTS CAN MEASURE APTITUDE AND ACHIEVEMENT, CAN DIAGNOSE, AND, MOST IMPORTANT, CAN BE AN EFFECTIVE TEACHING TOOL. MANY TEACHING TECHNIQUES ARE EXCELLENT FOR TESTING AND SOME TESTING TECHNIQUES ARE GOOD FOR TEACHING. WELL-PREPARED TESTS BEAR OUT WHAT A GOOD TEACHER ALREADY KNOWS, AND ALSO HELP THE STUDENTS BY TEACHING NEW ITEMS BY IMITATION, INDUCTION, OR ANALOGY AS THEY TEST. BECAUSE LANGUAGE LEARNING ACHIEVEMENT IS CUMULATIVE, AND REQUIRES ANALYTICAL POWERS PLUS GREAT AMOUNTS OF MEMORIZATION, HABITUAL REPETITION, AND PRACTICE, TESTS NOT ONLY SHOULD REQUIRE KNOWLEDGE OF NEW WORK BUT ALSO SHOULD BE COMPREHENSIVE ENOUGH TO SHOW THE STUDENTS' ASSIMILATION OF PREVIOUSLY LEARNED MATERIAL. TESTING IS ALSO A VERY EFFECTIVE MEANS OF MOTIVATING STUDENTS TOWARD SPECIFIC OBJECTIVES AND TERMINAL BEHAVIOR--PRETEST STUDY AND THE TESTING PERIOD ITSELF RECEIVE A STUDENT'S BEST CONCENTRATION. THE PRINCIPLE OF SPACED LEARNING APPLIES WELL TO SPACED TESTING. FREQUENT SHORT TESTS ARE MORE RELIABLE, PRODUCTIVE, AND FAIR TO STUDENTS THAN INFREQUENT LONG ONES, BUT BOTH SHOULD BE USED. SHORT SPOTCHECK TESTS ARE RELIABLE IF THERE ARE ENOUGH RECURRENCES OF ITEMS FROM A COMMON CATEGORY. ORAL WORK CAN BE GRADED EASILY BY USING A SIMPLE TAPED TEST. THIS SPEECH WAS DELIVERED AT THE MEETING OF THE AMERICAN ASSOCIATION OF TEACHERS OF SPANISH AND PORTUGUESE, CHICAGO, DECEMBER 27, 1967. (AS)

ED016476

A CLOSE LOOK AT THE IMPORTANCE AND PLACE OF TESTING IN THE FOREIGN LANGUAGE PROGRAM

Richard Barrutia
University of California, Irvine

The Functions of Testing. Testing of knowledge and skills can follow many procedures and serve many functions. In a foreign language program, however, testing takes on an even greater importance with relationship to teaching than it does in other disciplines. In language teaching these two activities sometimes merge to the point of indistinguishability, because the very nature of audio-lingual teaching requires such a great amount of feedback, reinforcement, and verification. Just as procedures merge, so do functions and results. For instance, at what point does a test begin to serve as (1) a measure of the predictability of language aptitude, as (2) a measure of the achievement of practiced skills, as (3) a diagnostic tool, or as (4) a teaching tool?

Aptitude. An early application of testing would be to determine, if possible, the student's level of language aptitude. This type of test gets deeply involved with linguistics, psychology, perception, imitative ability, cognition, and retention. The foreign language teacher is not expected to prepare this kind of test. The combined complexities of all these matters get too involved for us to deal with at length in this handbook or, for that matter, to be completely studied in depth by the average teacher. Suffice to mention that tests measuring predictability of language aptitude are available.[1]

---

[1] John B. Carroll and Stantly M. Sapon, Modern Language Aptitude Test, Form A (New York: The Psychological Corporation, 1958, 1959).

Pimsleur Language Aptitude Battery (New York: Harcourt, Brace & World, Inc.)

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Achievement. One of the most common functio that tests can serve is in measuring the achievement of previously-practiced skills. Here testing becomes immediately meshed with teaching, because in procedure it involves another repetition or practice session in spoken or written forms. This type of test is thus establishing for the teacher another example of the student's performance and serving the overlapping purpose of a teaching tool for additional practice.

Diagnostic. There remain two other major test types mentioned above, which play a great part in the ongoing teaching activities of a foreign language program. These are the test as a diagnostic tool and the test as a teaching tool. It would seem at first glance that the test as a diagnostic tool would be the same as testing for the predictability of language aptitude. This is not so, because the diagnostic test serves a continuing every day function and is helpful to the teacher throughout the program. What is meant here is that the teacher can diagnose, by the results of the test, what it is that might be going wrong either with his teaching or with certain inadequacies of the teaching materials. In so diagnosing a specific problem with which the class may be wrestling, the teacher can place more emphasis on the illumination of that problem, then proceed again in a regular way. Another diagnosis, which can be made, might well be one of an individual student's difficulty. In this case, of course, the teacher should then assign extra study, laboratory drill, or give individualized attention. The diagnostic test, therefore, also overlaps into the area of the test as a teaching tool.

Teaching Tests. The tests which serve more purely as teaching tools, however, are the ones that challenge the students to learn by analogy and

discovery and lead them to try to understand new items as the test proceeds. These tests check not only what the student has learned, but also test his ability as a continuing language learner. They help the teacher discover if the student is learning how to learn. Some items of this type in each test are often the most helpful to the student and teacher alike in the daily ongoing function of the foreign language probram.

Collateral Problems and Benefits. It is important to remember that good testing always accomplishes many things. Classroom tests are used to provide a basis for assigning grades, to compare pupils, to motivate learning, to direct learning, and to help improve the instructional process through measures of pupil growth. Classroom tests also provide pupils with their best source of information about instructional objectives and their best "feedback" concerning how well they have learned what they have been taught. If the tests represent faithfully the major course objectives, preparing for the tests and reviewing the tests after they have been administered will reinforce the other teaching that has been done.

Among other benefits, testing verifies the evaluation which a good teacher, who has done his work well, has already made of his students. When the test does not reinforce the teacher's opinion, several things could be amiss. Perhaps the test was a bad one. This should show up if there is consistency in a pattern of several good students doing poorly on it and vice versa. Perhaps the teacher's previous appraisal was not based on enough information, in which case the teacher is at fault for having made too hasty a judgment. A related problem might be that the type of test used, though a good one, might not have been used enough previously for the students to have become familiar with it. Many times a teacher's

evaluation of a student is quite correct, as it is probably based on the daily contact of classroom performance. But when there is insufficient testing to give the measured results necessary for the sake of the records, a single test may not corroborate the previous ongoing and more just appraisal.

The questions that are used in the classroom test provide pupils with better information about the teacher's course aims than do the textbooks or the course of study. They also direct pupil efforts towards the acquisition of the facts and skills needed for achieving these aims, thereby improving their performance on the next test. Perhaps most important of all, when the teacher sets his test specifications and tries to devise test questions that will help him in judging pupil progress toward particular educational goals, he begins to see more clearly what the goals really mean and how hard it is to measure pupil progress toward them.

Focused Attention. It is entirely possible that testing, correctly used, can become the most effective teaching tool ever devised since the time of Socrates. After all, is not the Socratic method merely a way of teaching by asking questions? All teachers should stop to consider why learning via a test is so much more efficient and long lasting than learning via other means, such as a test, or television, or the language laboratory. When we place ourselves in the student's shoes, the answer becomes quite obvious. Learning during a test is effective because more and specific attention is directly focused on every item of a test, especially when the student knows that every item is going to be counted. This is precisely why programed instruction has made such dramatic inroads into teaching methodology in all areas of education. A programed course is built up of a series of hundreds of small tests. The student is required by the program

to react correctly at each step before he proceeds. When he is right, he gets a positive reinforcement, then a new item. When he errs, he is forced to reflect on the error immediately. He is then reinforced positively when the same item appears again later in the program. The reliability is vastly improved by frequent testing with immediate reinforcement.

To sum up this notion, then, testing is one of the most effective teaching tools simply because it is a review, which is given added prestige and importance by being a checkpoint labeled as a quiz, test, or examination and, consequently, honored with more attention by the learner.

Spaced Learning. Another notion, which has not had enough consideration in the past by foreign language teachers, is that of spaced learning. This principle holds that in foreign language programs it is better to teach or drill a given segment for two minutes, ten different times, than it is to teach it for a twenty-minute period at one time. The reader might well ask: So how does this principle relate to testing? In testing, it is safe to stipulate that more frequent short tests are more effective than infrequent long tests. Various beneficial results can readily be seen in following such a practice. The students become used to the different types of items that might appear on a final exam. This contingency makes the larger exam more just and more reliable because the students are now comfortable with the format. By employing frequent exams, the teacher might diagnose problems which he can correct long before the midterm or final exam comes along. Frequent exams assure the teacher of more occasions of heightened concentration and attention on items that otherwise might have been just another repetition or review. And, finally, frequent exams give still another opportunity to teach by analogy--the teacher's most helpful ally.

Some Specific Techniques. It seems to develop quite naturally, then, that in language teaching, more than in most disciplines, testing is teaching and vice versa. For example, suppose you as a teacher wish to check if your students can make logical and well-chosen variations on the dialogues they memorize. A simple way is to put a series of phrases on the chalk board from which they should select in order to alter their dialogue. They would then orally engage their partners in the dialogue, making substitutions of whole phrases selected from those on the blackboard. Their dialogue partners are thus obligated to adjust to the variation and, in turn, pick a substitution from the board that semantically fits their counterpart's conversation. In this little test, the teacher is not only checking for ability in dialogue variation but also checking reading aloud, reciting from memory, and comprehension. It can now be seen how this test teaches just as well and as much as it evaluates, as do most good exams.

Impromptu Testing. In purely oral testing, too many teachers seem to feel that they are restricted to entirely structured formats, just because it is a test. This is far from true. In fact, the seemingly unstructured reality of certain tests are often their best quality. One good obvious way is for the teacher to simply move about the room quickly, asking as many rapid fire questions as can be crammed in to an active ten minutes. These questions can be about anything but staying, of course, within the student's range of vocabulary and grammatical understanding. One slight problem, which arises with this procedure, is the matter of lack of time for grading the student responses. This difficulty is resolved with a pretaped test of nearly this same type, which will be described later.

Directed Testing. Another type of loosely-structured oral test is an outcropping of our classroom directed dialogue technique. The teacher instructs Student One to tell Student Two to ask Student Three if he or she has done or will do something or other for him (Student One). It is important that in responding, all three students become verbally engaged in the exchange using complete sentences. This forces them all to modify the answers and inflect the verbs in various ways. The objection might be raised at this point that these are nothing more than audio-lingual teaching techniques, but it must be clear by now that techniques also merge into overlapping areas of teaching and testing.

A New Tape Technique. The two examples of the preceding paragraph both have the disadvantage of being difficult to judge and grade, because this happens so quickly and because after the student's utterances, there is nothing left for the teacher to rely on for assigning a grade but an overtaxed memory. Perhaps the following arrangement will overcome this disadvantage. Prerecord many questions and statements that take rejoinders. Leave a pause after each statement long enough for a student to answer, then record a random number followed by another pause. Use as many numbers as there are students in the class, and repeat each number as many times as necessary to assign a number to every question recorded. Assign numbers to the students especially for the test (have them count off in the target language). Play the tape, and have the student whose number comes up answer orally first, and then write it out on the test sheet. The students will all be alert to answer, since the numbers are recorded at random. Anticipating their number, they will mentally formulate the answers to all

of the questions, even though each student may not have to answer and write out more than five or six. This system frees the teacher to move about with the grade book in hand and allows enough time to make a judgment and to give a grade on oral production. It also gives each student plenty of time to write out his answers, because the same number is never recorded twice in succession. These small test papers are collected and can then serve the teacher at his leisure to decide on an additional grade for written production.

Pattern Paragraph Practice. When students have become familiarized with certain teaching techniques, they are better able to handle tests couched in more or less the same format as the regular classroom presentation of the material being tested. What is more, a test of this type has much higher reliability than one which presents the same material in a format that the student has never encountered. This principle holds true for written work as well as for oral production. A clear example occurs with "pattern paragraph practice." Suppose the class has been practicing paragraph writing by substituting the lexical items and leaving the function words intact, thereby adhering to the original grammatical structure of the paragraph but altering its semantic sense. This is an excellent writing practice that works well, the successful results of which can easily be confirmed. In classroom practice the instructions could be as follows: "Copy the following short paragraph first as it is. Then copy it again changing the underlined words to words of your own choice that make sense and are grammatically correct within the reconstructed whole." On a test, practically the same instructions will serve well: "Change each underlined word to any word of your own choice." Teaching and testing of this

combined nature is the most reliable, logical, just, and beneficial for all concerned.

Directing Motivation. Another purpose served by good testing in addition to verifying assimilation of the material covered is motivation toward certain objectives and a specific terminal behavior. This purpose is all too often not even recognized by many teachers. Let us try here to elucidate and expand upon it. It was mentioned above that at the moment of testing, the student is at the peak of his powers of attention and concentration. Therefore, we suggested that the teacher should take advantage of this peak performance and include some test items, which are clearly designed to make the student try to see relationships which he perhaps had not seen before. In other words, it's a good time to teach. Now taking this point of advantage a step further, we come to the student's study time when he is preparing for a test he knows he is about to have. Much of a student's best study effort comes when he is preparing for a quiz or an examination. Furthermore, this heightened and sharpened study effort can be directly focused on any given set of elements the teacher cares to emphasize. All the teacher need do is to be sure that the students know what to expect in terms of the material which will be covered, and that they be familiar with the technique that will be used. Testing thus becomes the strongest motivating force available to the teacher, and it is a motivating force which can be directed toward certain objectives at that. The natural conclusion, then, of this particular point on testing is obvious. If we want to maximize extended effort on the part of our students at home and in class, we should plan to test frequently. Short, frequent tests concentrating on one or two skills at a time will have the greatest

effectiveness.

Generalizing. We refer to the generalization that is so often necessary in evaluating tests, which merely sample or spot check. Not knowing exactly which items of, let's say, the regular past subjunctive verbs will appear on the test, the student is apt to study that whole category, which is exactly what we want him to do. But in correcting a short test of this type, the reliability factor might be questioned if the whole test is too short of if there are not enough items (approximately three) of each verb type included. Pairs of items permit more reliable judgments than single items. Triads of items would be still more reliable than pairs. For the most part, however, it is safe to assume that if a given student can properly conjugate three regular -ar verbs, he can conjugate six and that if a student can write two correct hypothetical sentences contrary to fact, he can write another. We may, with caution, assume that this principle can be extended so far as to say that if a student can write an acceptable paragraph, he might also write an acceptable composition. There is still another phase to this trend of thinking, which is that the consistency of production of a given group of students also allows us to generalize about the rank order of the group. The top student of the class on one day is not apt to be at the bottom on the next. Though the weakest student may have at some point a shining hour, his overall performance is sure to betray him.

Listening Comprehension. The one language skill which is probably least understood yet has the greatest potential for total learning is listening comprehension. We have often heard people say, "Oh, I can't speak your language very well, but I understand everything." The only

situation where such a case might be true is when the speaker knows that his listener is indeed not very proficient, so he speaks slowly and carefully trying to use the vocabulary that the listener does command. But the moment that that same learner, who "understands everything," is among native speakers of the target language who are conversing at their normal rate of speed and in their normal way, he is lost. Comprehension must be worked on, then, early and constantly, and it must be tested at regular frequent intervals. The importance of this phase of foreign language learning cannot be over emphasized. Good understanding of a foreign language as it is normally spoken by natives concerns recognition of all sandhi (liaison, synalepha, and hiatus that cause forms to change), variation patterns, all phonological and grammatical interferences, all intonation differences, and all cultural implications. When we add the problem of reducing this awesome list to the writing system of our target language with all its lack of "fit," the task becomes formidable. Nevertheless, all these items can be practiced in comprehension assignments, and they can be well tested by following the same techniques.

Listening-in Procedures. Comprehension practice and tests can be conducted by having the student listen to recorded material with which he is unfamiliar either in the language lab, in isolated booths, or at home if he has the equipment. His task is to write out the script of a given lesson, taking as much time as is necessary to transcribe it all as correctly as he can. In the classroom, the teacher has several students copy certain sections on the blackboard in order to have the entire script. While this is taking place, the teacher should ask other students to read certain sections of their assignment. The whole piece is then finally carefully studied for meaning, grammatical difficulties, spelling, and

sound by reading it aloud. This practice, though not a new one, is too often overlooked in spite of the considerable improvement each session brings about in the student's listening comprehension.

Now in a testing situation, almost the same procedure obtains. The only difference would be the follow through. That is, in testing we tend to mark down the errors rather than correct them on the spot. Nevertheless, the comprehension test can be a prerecorded paragraph, dialogue, or narration with which the student is expected to deal in almost the same manner as if it were a practice session. He would be limited, of course, only to the amount of time he could spend on writing out what he hears. Looking at the test in this way, it becomes very much like a dictation. The teacher may, in fact, prefer to read the passage himself, doing it several times as if it were a dictation. Because of the tremendous importance of ear training for complete understanding, may we repeat that the comprehension skill is the most important acquisition a teacher can give the students. Not only does it pay off in kind by increasing all the other skills, but the pay-off to the teacher will come every time the students gratefully acknowledge that they are understanding the foreign language movies, radio broadcasts, television shows, and their native-speaking acquaintances. The teaching practice described above is, or should be, the nearest approximation possible of an individual transported to the foreign country. And the added advantage is the availability of closely supervised instruction. As a testing technique, it is extremely effective for testing any or all of the five skills of understanding, speaking, reading, writing, and cultural assimilation.

A Note of Caution. It is not the intent of this chapter to lead

the teacher to believe that teaching and testing are so much alike that all testing techniques should or can be used for teaching. While it is mainly true that most teaching techniques can almost always be adapted for testing, the reverse is not always so true. For instance, a good test ideally should test everyone in the class. In doing so, it must challenge and indeed go slightly beyond the competency of the best student in the class; otherwise, the limit of that student's ability still remains an unknown. This means that if there is even one perfect paper on a given test, strictly speaking, that test has not measured the whole class. Now that single student, who has not had his limits challenged, does not invalidate the test as a good average measure. That is why we say that a test should ideally measure the upper limits of everyone. As a teaching procedure, this notion is often effectively employed, but it is dangerous to always aim over the head of the best student. To repeat an old cliché, "Nothing succeeds like success." The class needs positive reinforcement continually, or it will lose interest and momentum. At the other end of the testing spectrum, if a test item is so easy that every student answers it correctly, such an item cannot discriminate between students. Hence, strictly speaking, it may be considered as an invalid item. In teaching, on the other hand, we do this every day in class. That is, we give out challenges to which the entire class can respond successfully. In so doing, we create an effective drill. One purpose of structure drills is to internalize habits by many successful repetitions. Here, then, is a marked difference where a teaching tool can serve for testing, yet a testing format does not suffice as a teaching procedure.

Re-entry of Material. It is important for the entire class to know

precisely what material each test will cover and what skill will be emphasized. But the students should realize that acquisition of a second language must be accumulative and that grammatical points covered earlier in the course will continue to be involved in successive tests. So that the student can be checked on his ability to transfer his knowledge to new problems, there should be tests recombining a number of points at frequent intervals. Some students, who do well when the class is working on certain elements, may be at a loss when called upon to produce those same structures spontaneously at a later time. When this occurs on a test, that test is serving a diagnostic purpose and should act as a signal to the teacher to use some re-entry and recombining teaching techniques at the very first opportunity.

Tests fail to provide a valid appraisal of the students' assimilation of the new material when they are concerned only with keeping abreast of the text and the points studied most recently. The most effective tests are those that vary somewhat in their techniques. But, more importantly, they must be well balanced and comprehensive in the coverage of material from any given point in the foreign language program.

Differences. It might be prudent here before winding up this chapter to point out that more has been said here about how and when teaching and testing merge than how and when they differ, which must also be an important consideration. The main difference would be that a test should always concentrate more on what to test rather than how to test. Discussions of testing should not deal only with techniques as the chief concern. We are on the right track when we ask ourselves such questions as, "Does the strategy that we use adequately measure the exact problem in the specific

skill which we wish to test?" Another caution to take would be against a misunderstanding of the test, which serves so heavily as a teaching tool, that it may very well be measuring a student's native intelligence and language aptitude rather than his achievement. The only thing that is radically different about teaching and testing is what we do with and about the results. In a teaching situation, we carry on the activity to sometimes profit from the student's error, while in a testing activity we must stop to make a note of how much has been assimilated. We should carefully define the problems that we want to test, then be sure that we observe and score only those points. The other elements, which serve as a vehicle for the problems, then become our teaching bonus. The all pervasive nature of grammar and phonology in all forms of good language testing will guarantee the salubrious by-product of learning.

Conclusions. In summing up some of the answers and conclusions that we can arrive at here, we recapitulate certain salient points and find that:

G (a) Tests play several major roles in measuring (1) aptitude, (2) achievement, in (3) diagnosing, and (4) serving as a teaching tool.

O (b) If a good teacher has done his task properly, a good test will bear out what he already knows.

O (c) Exams can be of great help to the students when they also teach new items by imitation, by induction, or by analogy as they test.

D (d) Testing is one of the most effective teaching tools, because both study time preparing for a specific test and the testing period itself receive the student's best attention and concentration.

GOOD

T (e) The principle of spaced learning applies equally well to spaced testing, e.g., it is better to test for five minutes on ten different occasions than to test for fifty minutes at once.

E (f) When a test does not verify what the teacher thinks he knows about a class or a student, it then serves as a diagnostic tool more helpful to the teacher than the learner.

S (g) Many teaching techniques are excellent for testing.

T (h) Testing is teaching and vice versa, but testing practices are only slightly less useful for teaching than are teaching procedures for testing.

S (i) Short, sample, spot-check tests are reliable as long as there

TESTS are enough (approximately three) recurrences of items from a common category.

T (j) Frequent, short tests are more reliable, more productive, and more fair to the students than infrequent long ones. But both long and short tests should be used.

E (k) Comprehension practice is often neglected, yet it is most effective for testing all the skills.

A (l) Motivation toward specific objectives and terminal behavior is more attainable by testing than by many other means.

C (m) Giving grades on oral production can be substantially aided by the use of a simple tape recorded test.

H (n) Foreign language learning and testing are different from most other disciplines in that achievement is accumulative and

TEACH requires analytical powers plus massive amounts of memorization, habitual repetition, and practice.

Traditionally, students are supposed to hate and fear examinations. But it has been shown that when the proper attitudes about testing are instilled in the students by a good teacher, this tradition can be reversed. Students are not averse to taking written tests that are lively and brief, nor oral tests that are bright and that seem spontaneous. Indeed, if some of the principles listed above are put into action, the teacher will take pride in his frequent tests. The students will study, learn, and be grateful for the opportunity to help program their own course and to learn Socratically. The test of a good teacher is to test his students.